



# GIS for CYP-C Research

Based on research by the lab of Dr. Alvaro Osornio-Vargas

Presented by Charlene Nielsen [[ccn@ualberta.ca](mailto:ccn@ualberta.ca)]



**UNIVERSITY OF ALBERTA**  
**FACULTY OF MEDICINE & DENTISTRY**  
Department of Pediatrics

# Outline

- Overview of GIS
- Basemaps and map projections
- Geolocation
- Analyses: statistics, overlay, and proximity
- Other learning references



# What a GIS is not

## GPS

Global Positioning System

GPS data can be used in GIS analyses

## Static Map

Digital/paper map is an “input” or “product” of GIS

A way to visualize output from GIS analyses

## Software

Functions and tools needed to store, analyze, and display geographic information

Requires hardware, data, and personnel in a complete system

## Database

Set of tables containing data that can be accessed or reassembled in many different ways

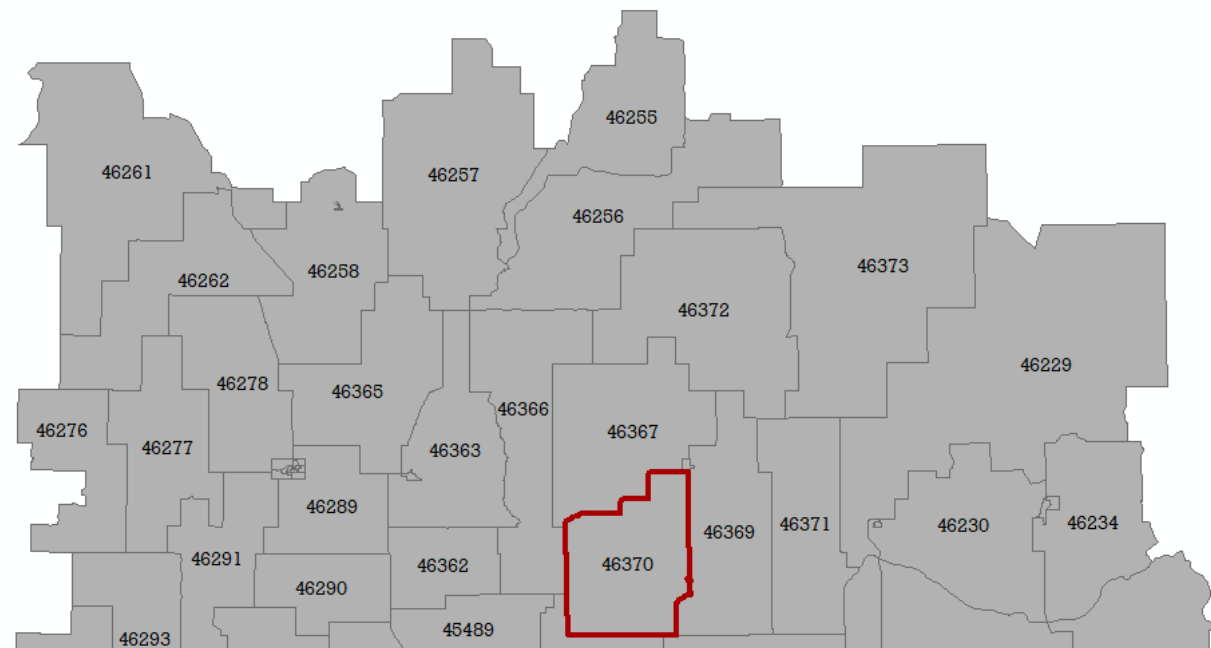
Requires a link to spatial data



# Geographical Information Systems



Link databases  
and maps



OBJECTID	Shape *	DAUID	CDUID	TotalPop	ChildrenPop	SES	Type	AREASQKM	ChildDensity
46367	Polygon	48130309	4813	478	135	0.55	2	288.73	0.47
46368	Polygon	48130310	4813	505	85	0.07	2	1.68	50.74
46369	Polygon	48130311	4813	486	115	-0.0	2	244.2	0.47
46370	Polygon	48130312	4813	498	100	0.17	2	229.75	0.44
46371	Polygon	48130313	4813	528	130	0.04	2	223.77	0.58
46372	Polygon	48130314	4813	504	130	-0.1	2	384.36	0.34

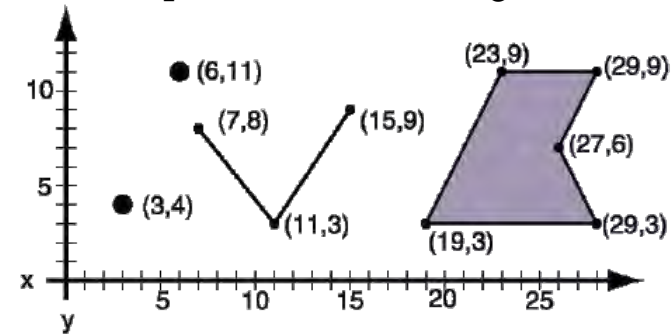
Dissemination Areas (1 out of 140 Selected)



# Types of spatial data

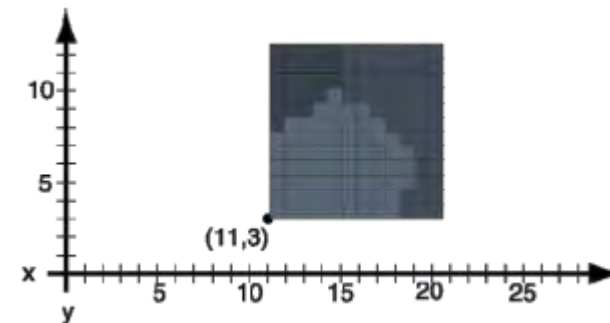
## Vector

Models discrete features as points (coordinates), lines (arcs), and polygons (areas) with precise boundaries and shapes with attributes; e.g. feature class, shapefile, CAD drawing

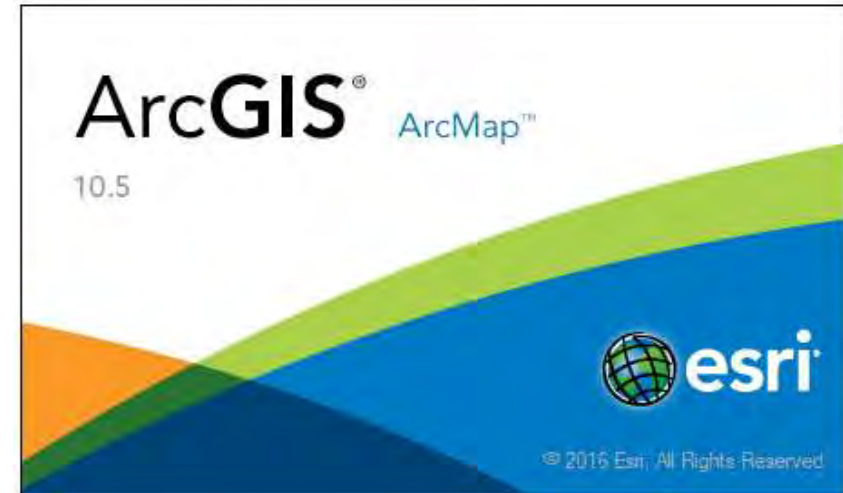


## Raster

Models continuous phenomena in a surface divided into a regular grid of cells (pixels) each having an associated attribute value; e.g. grid coverage, TIFF, digital photo



# ArcGIS Desktop



<http://www.esri.com/en/arcgis/products/arcgis-pro/Resources/ArcMap-Resources>

“Powerful desktop applications create maps, perform spatial analysis, and manage data in 2D and 3D.”

“ArcMap is the industry leading, traditional GIS authoring and editing application.”





Main Menu

Tools toolbar for working with map layers

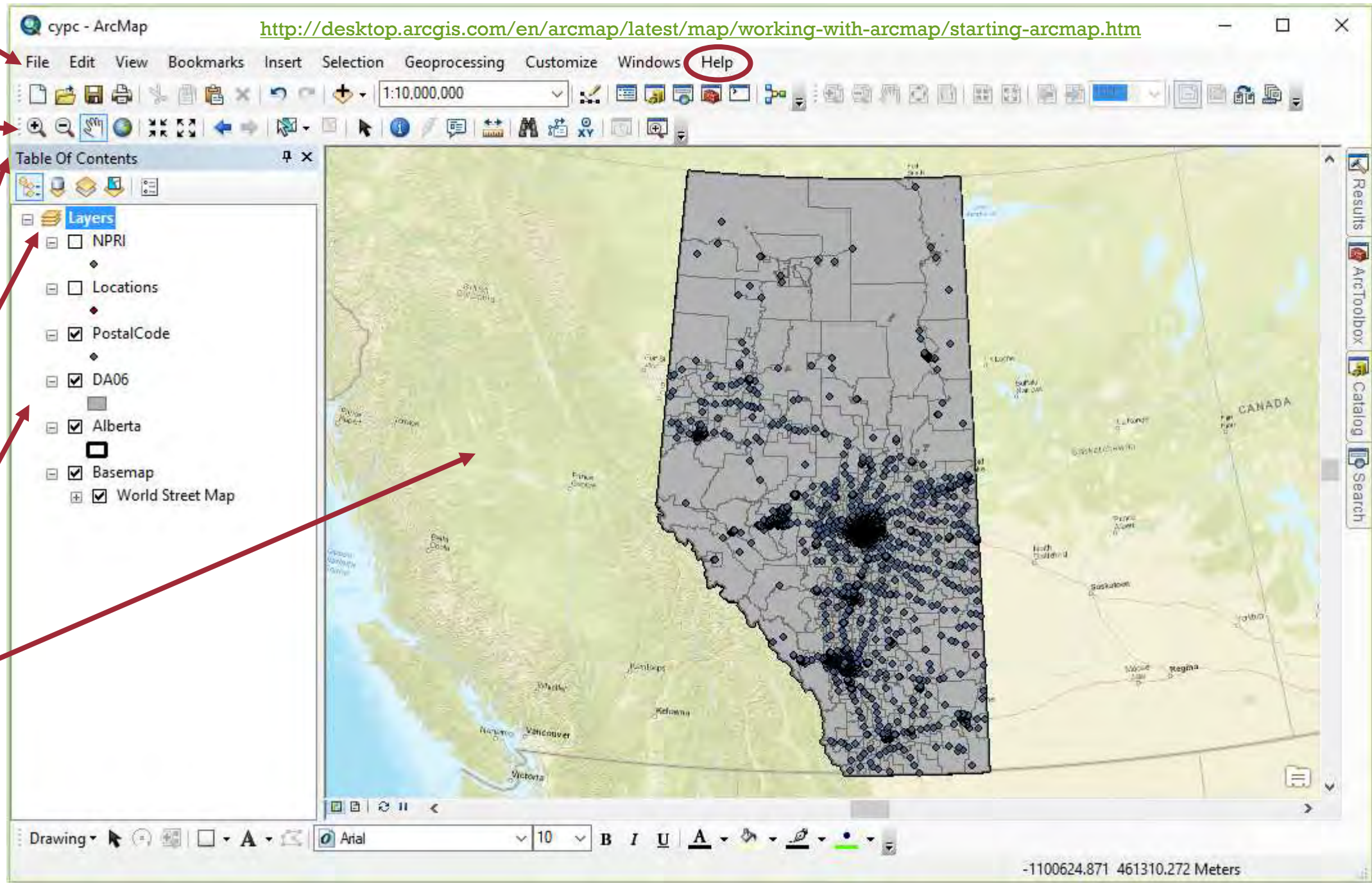
Table of Contents lists the items in the map

Data Frame contains the layers

List of map layers and their drawing order

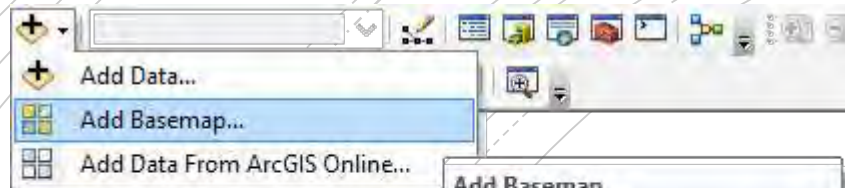
Data View

ArcMap



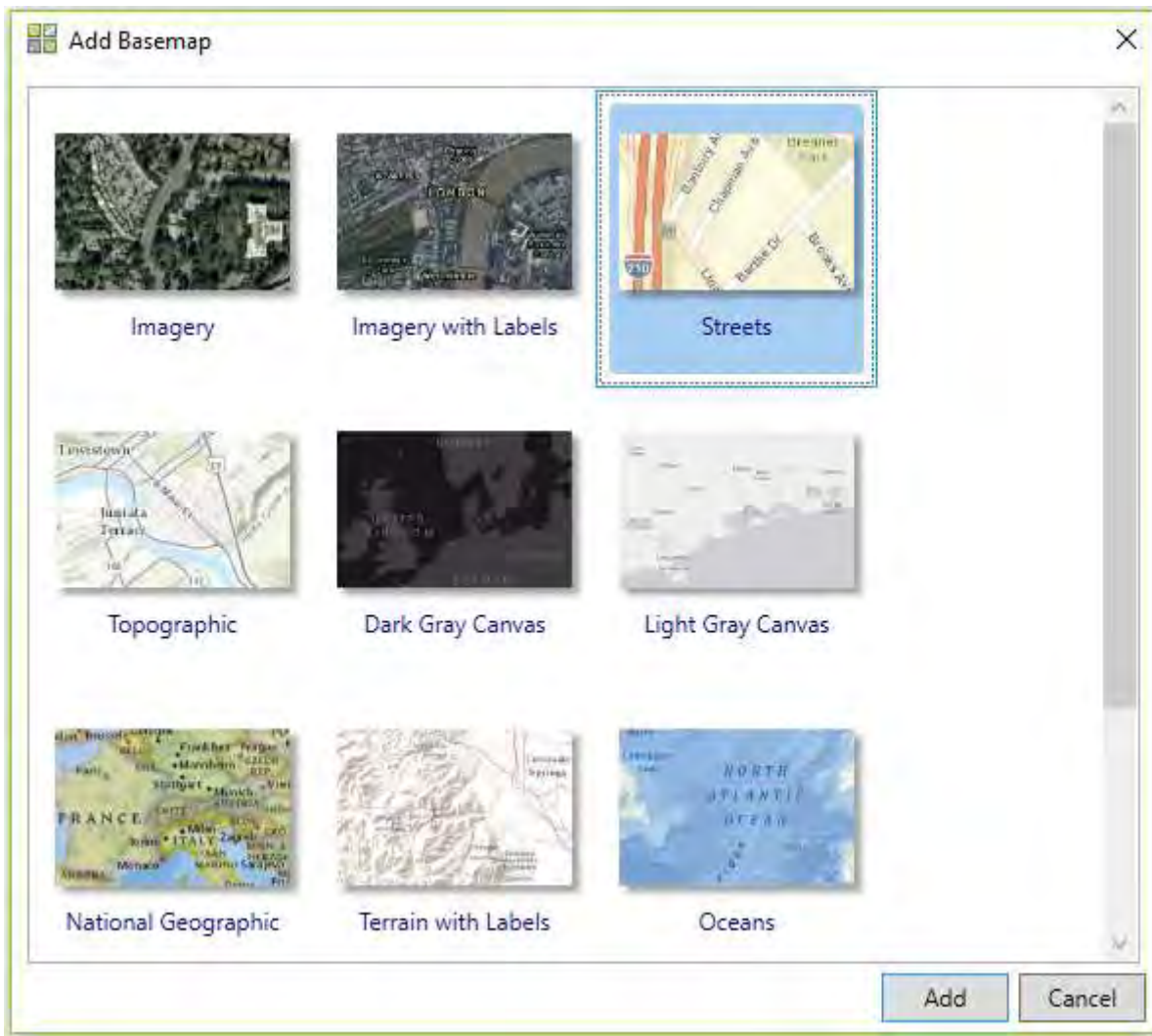
-1100624.871 461310.272 Meters

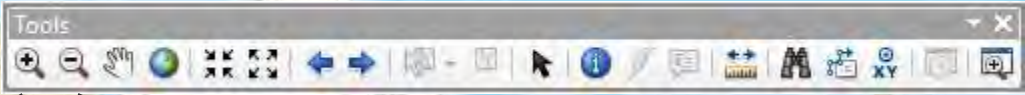




**Add Basemap**  
Choose a basemap from ArcGIS Online. These basemaps are map services that require an Internet connection for them to draw in your map.

# Basemaps





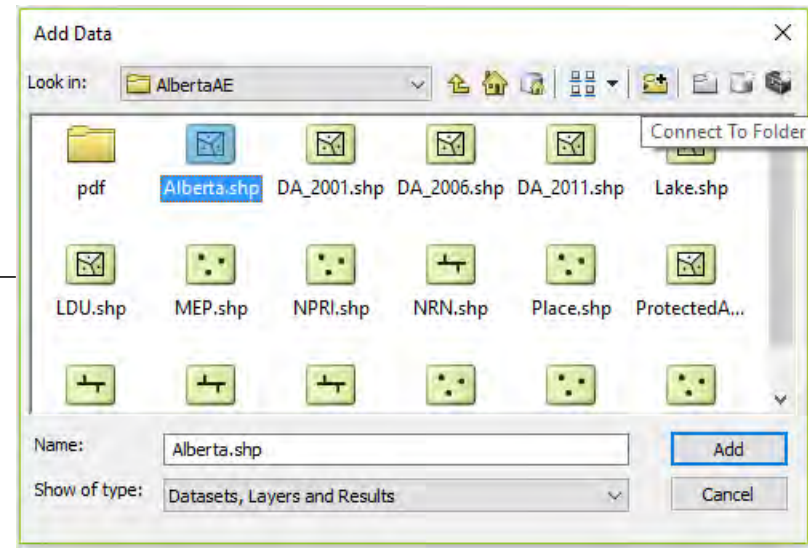
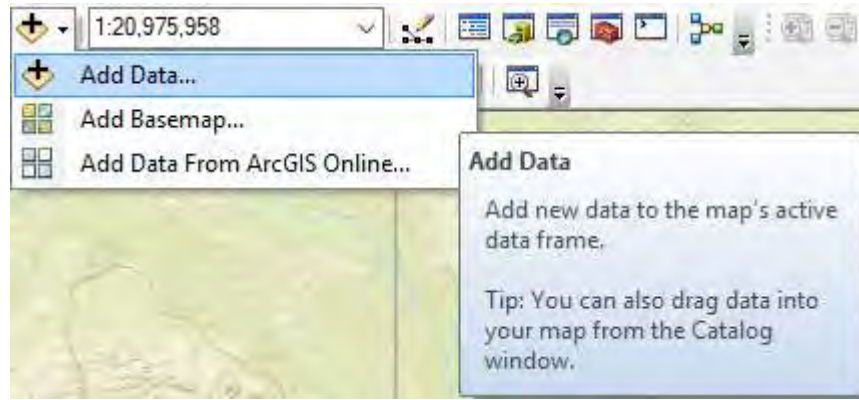
**Zoom In**  
Zoom in by clicking once or dragging a box on the map.  
  
Shortcut: Hold down the Z key, or hold down CTRL and drag with the middle mouse button held down.

**Zoom Out**  
Zoom out by clicking once or dragging a box on the map.  
  
Shortcut: Hold down the X key.

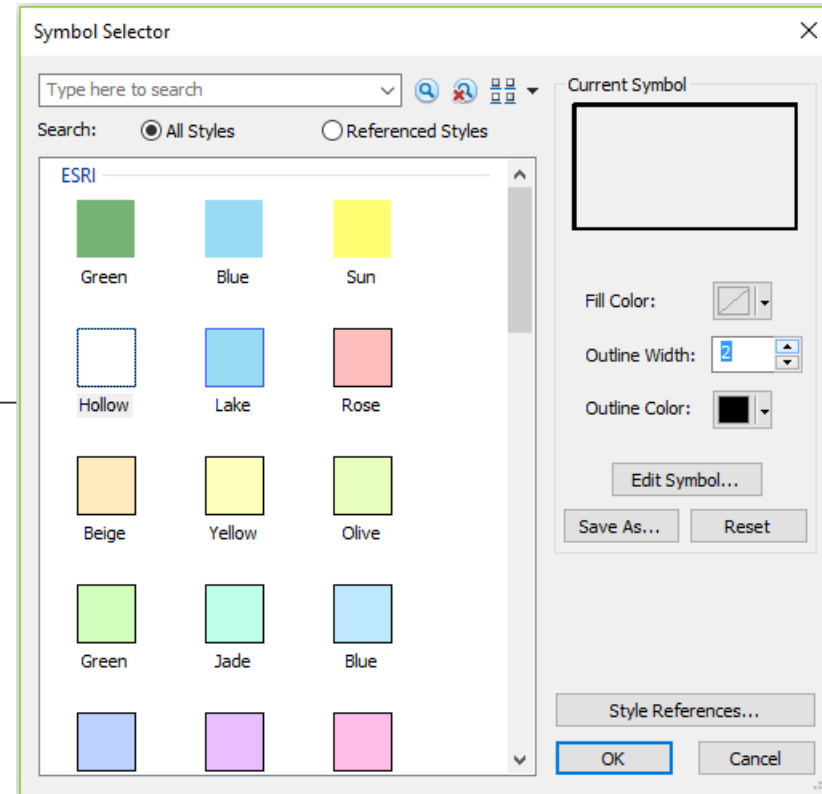
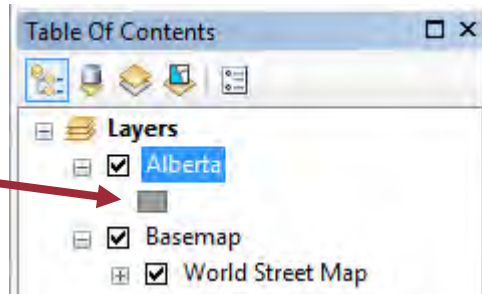


Navigate

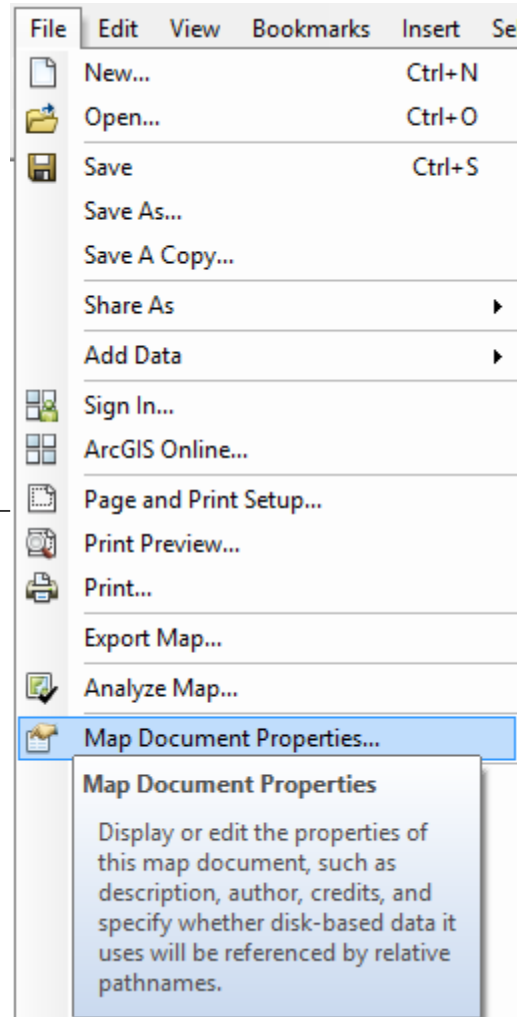
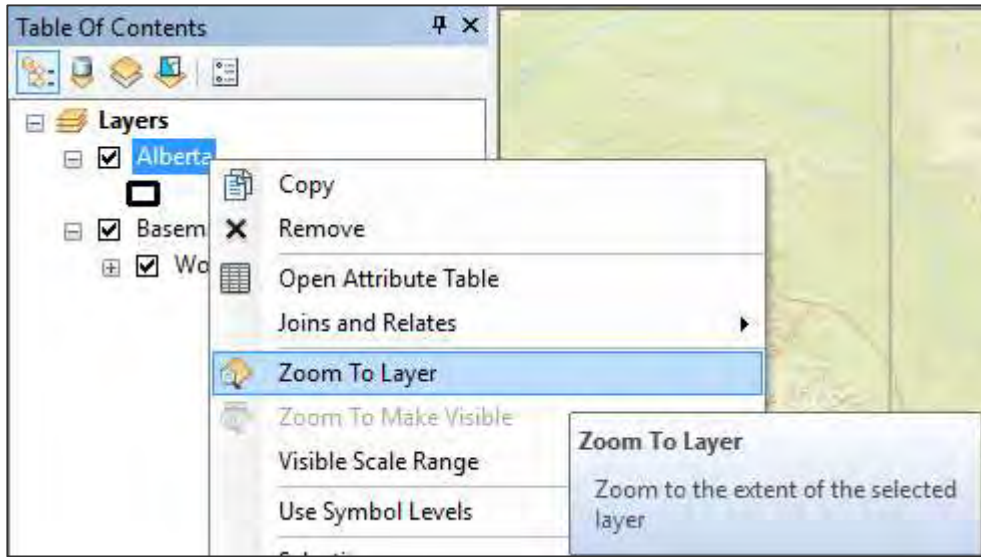




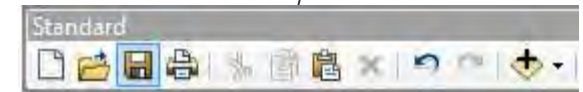
Symbol patch



Add data and symbolize



Pathnames:  Store relative pathnames to data sources



**Save (Ctrl+S)**  
Save the current map document.

Zoom and save

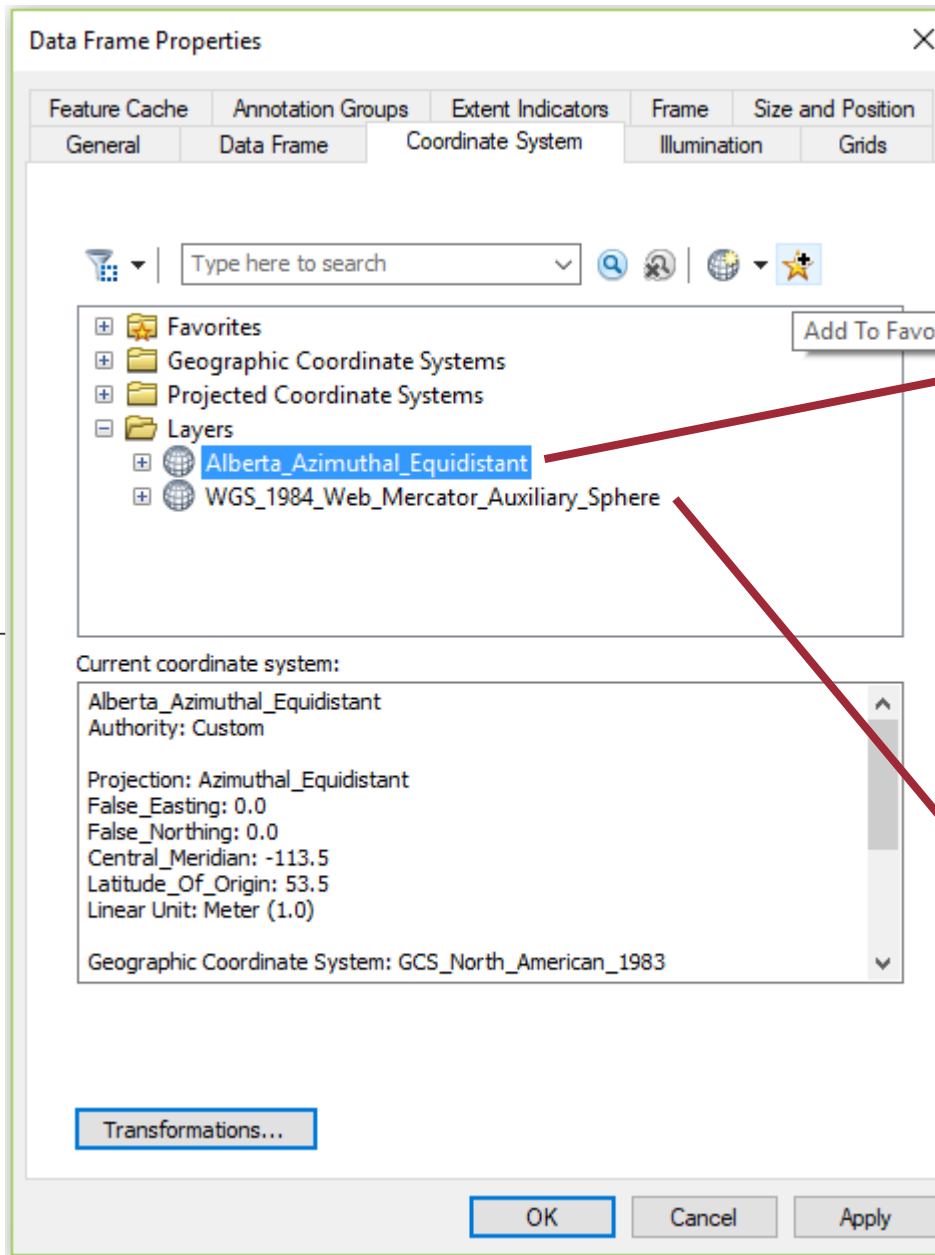
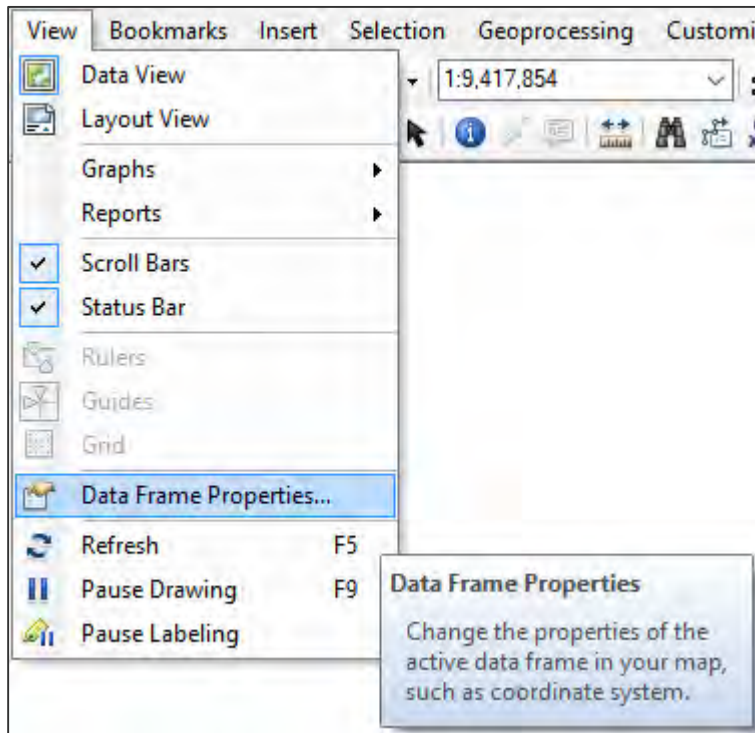
# Map projections

Name	False Easting	False Northing	Central Meridian	Standard Parallel 1	Standard Parallel 2	Latitude of Origin
NAD 1983 Canada Atlas Lambert	0	0	-95	49	77	49
Alberta Azimuthal Equidistant	0	0	-113.5	n/a	n/a	53.5
NAD 1983 10TM AEP Forest	500000	0	-115	n/a	n/a	0
NAD 1983 BC Environment Albers	1000000	0	-126	50	58.5	45
NAD 1983 Ontario MNR Lambert	930000	6430000	-85	44.5	53.5	0
NAD 1983 Quebec Albers	0	0	-68.5	46	60	44

<http://desktop.arcgis.com/en/arcmap/latest/map/projections/what-are-map-projections.htm>







Coordinate system

# Geolocation



## GIS Dictionary

Look up terms related to GIS operations, cartography, and Esri technology.

# A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

⊖ Back to list

<http://support.esri.com/en/other-resources/gis-dictionary>

### geolocation

- [geolocating] The process of creating geographic features from tabular data by matching the tabular data to a spatial location. An example of geolocation is creating point features from a table of x,y coordinates.



# Geolocation

Clean and format  
table of x,y  
coordinates

Add table to  
ArcGIS Desktop

Join by postal  
code to access  
longitude/latitude  
coordinates

Export joined  
table

Display XY Events

Export data to a  
spatial file format



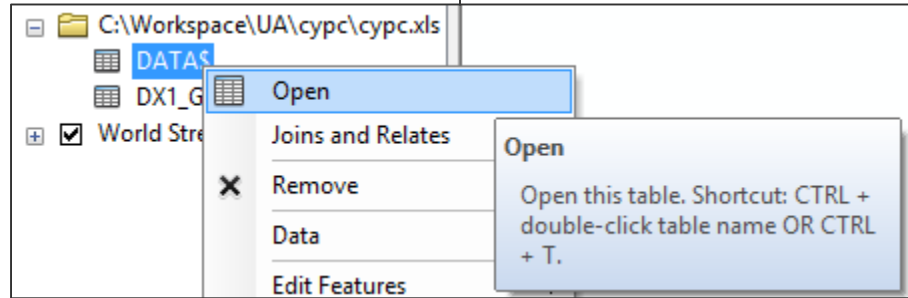
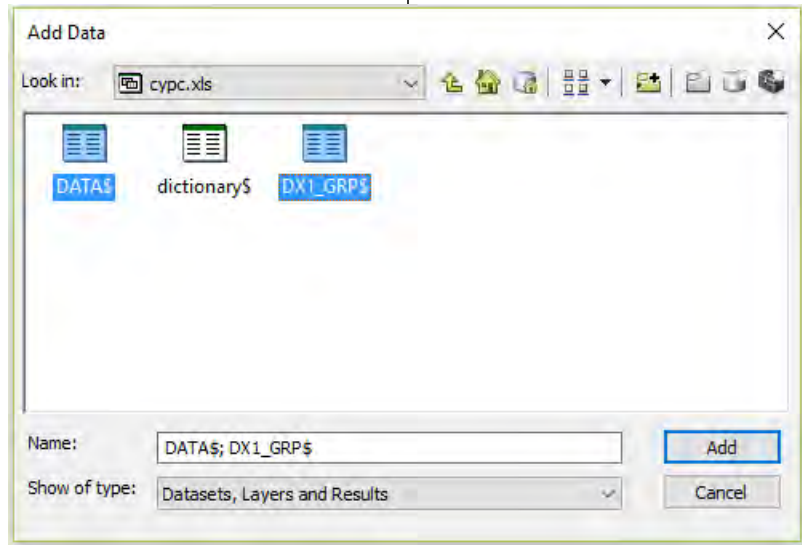
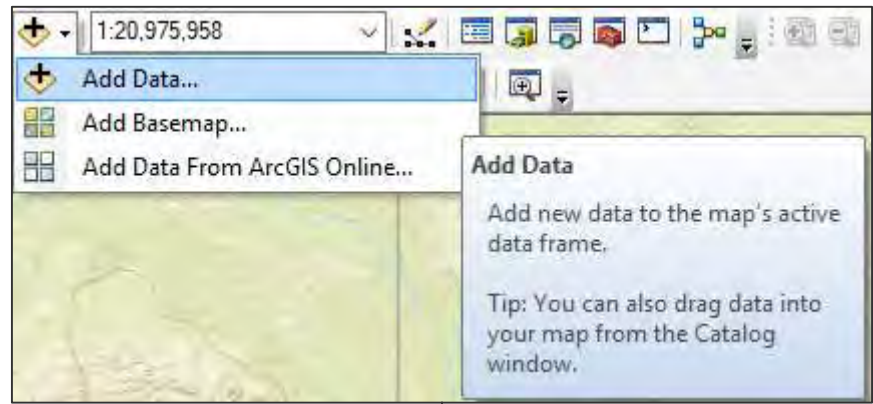
## Standardize the Attributes

- No database is initially ever “clean” and getting your data ready for spatial analysis will take upwards of 80% of your time
- Full 6-character postal code
  - CAPitalIZaTioN matters  
T6G 1c9 ≠ T6G 1C9
  - Ensure all are in the format of A1A 1A1  
T0L 0C0 ≠ T0L 0C0
  - Remove all spaces  
T6G1C9\_ ≠ T6G1C9

## Tips for Tables

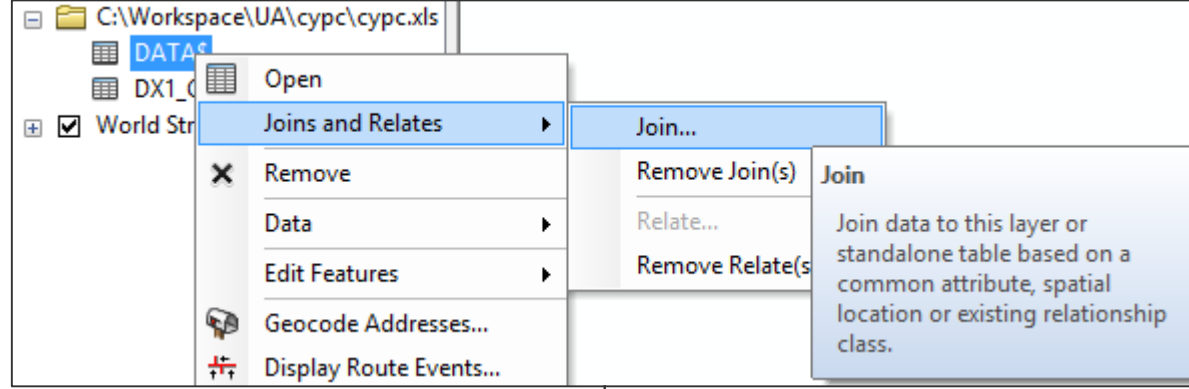
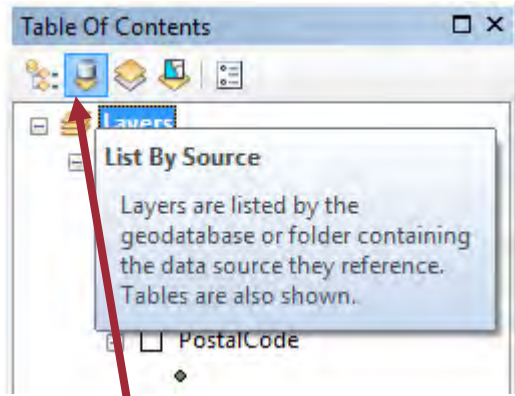
- Column headings must be present
  - No non-alphanumeric characters (e.g. %)
  - Do not use spaces – check at end of the word
- No skipped rows anywhere
- Date/time values are subject to import errors
  - May split the date parts in to separate columns
- Export table from your statistical software to:  
\*.dbf, \*.csv, or \*.xls
  - ArcGIS can only read MS Excel version 2003 or earlier (no \*.xlsx files)

Clean and format

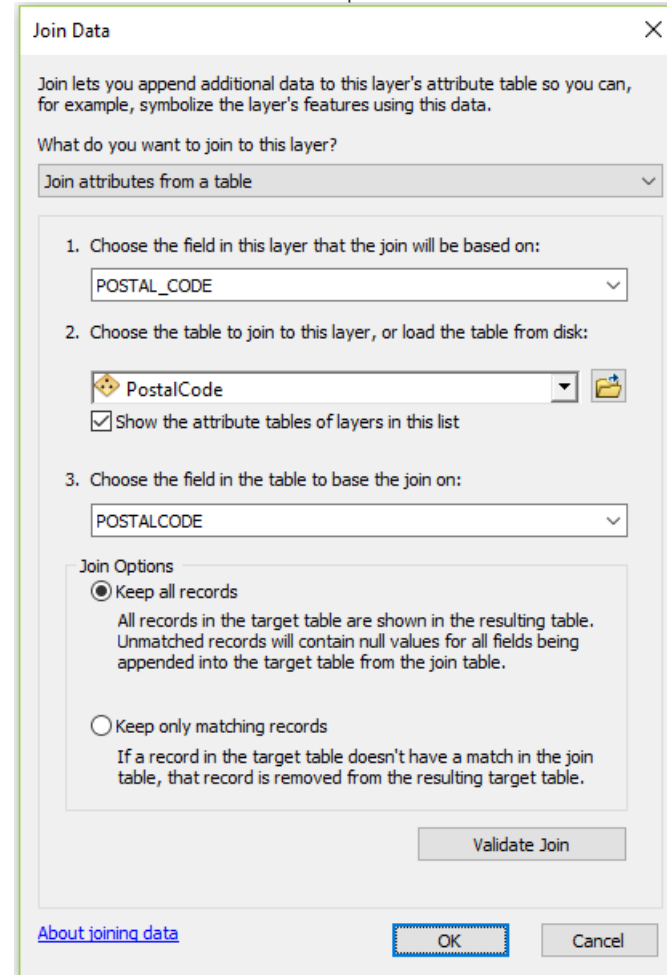
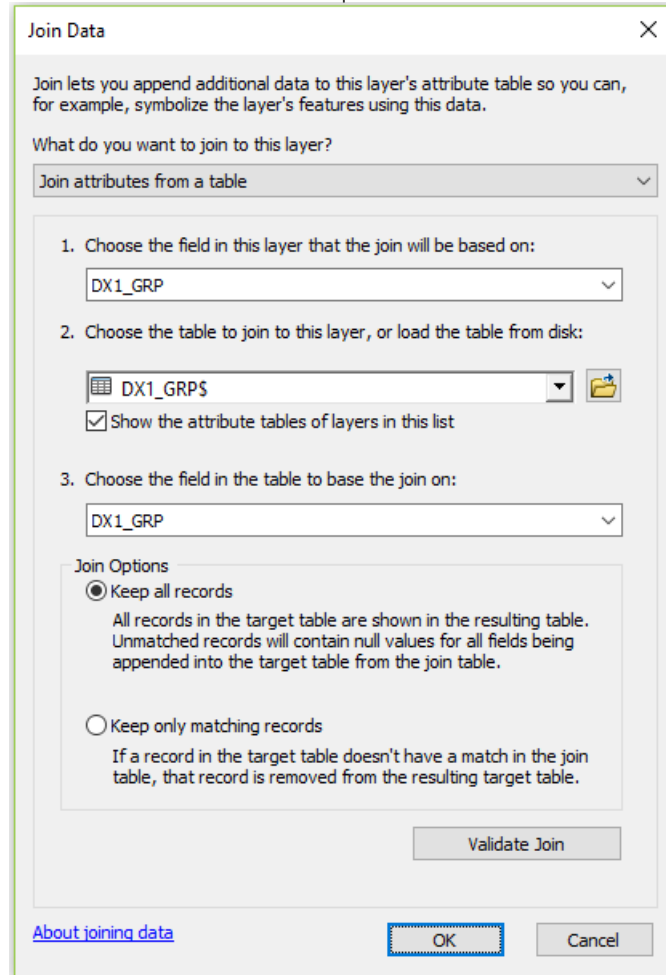


Add data



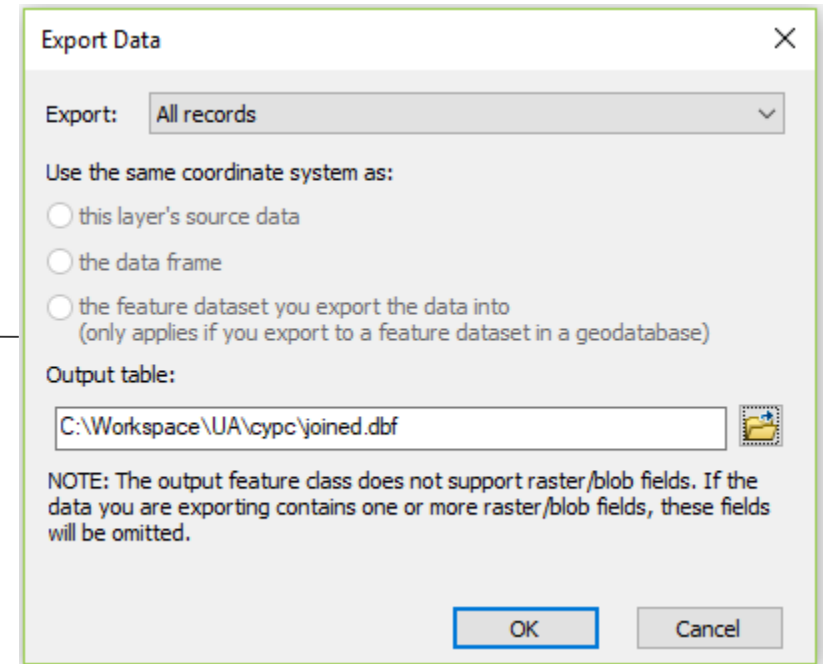
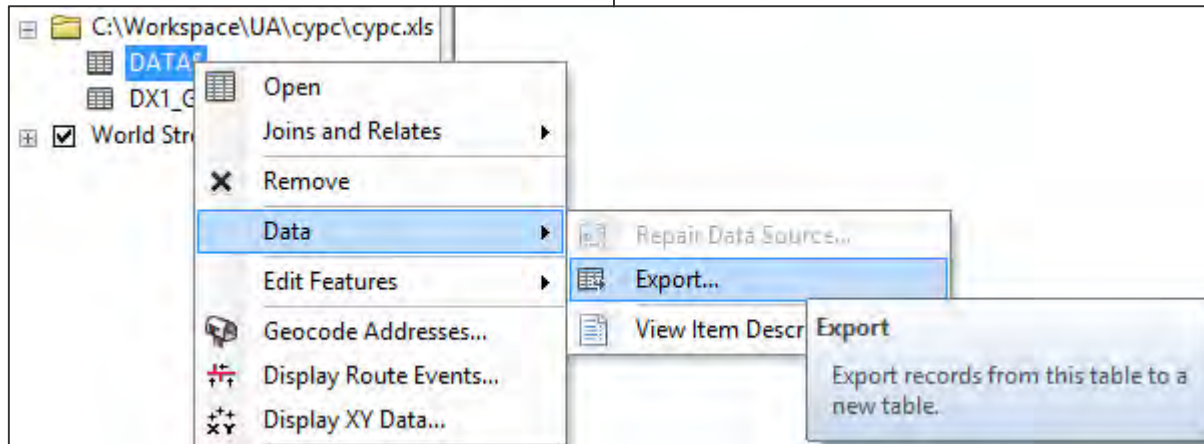


To see stand-alone tables in the Table of Contents, click the List By Source icon

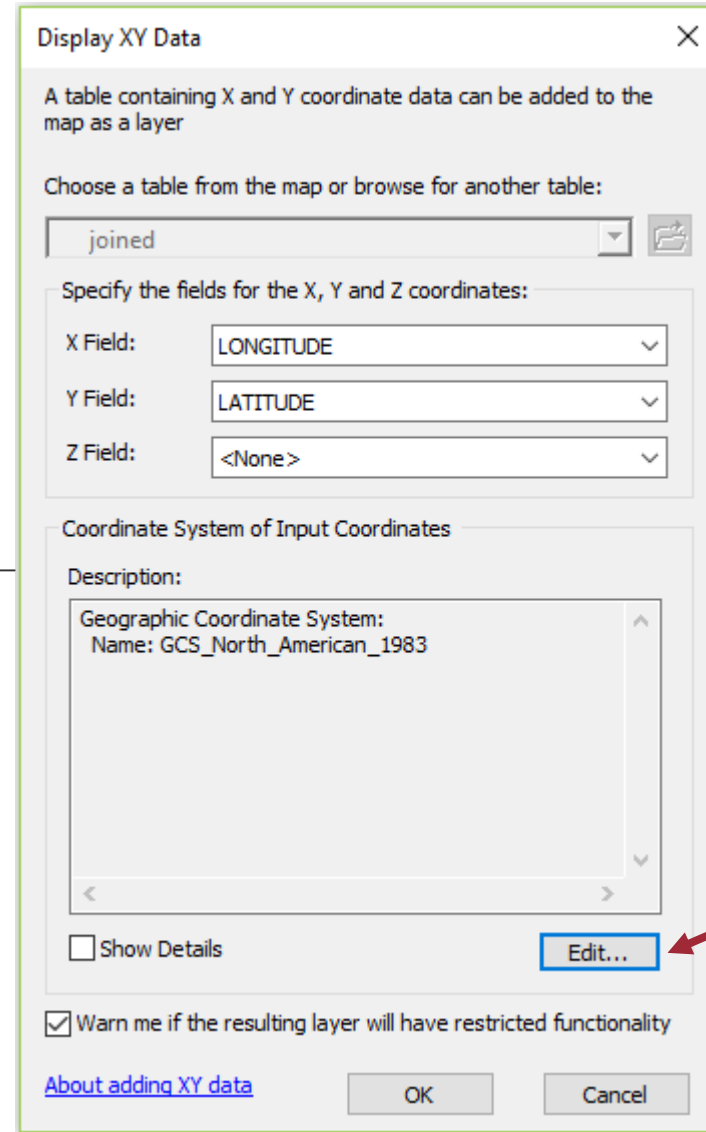
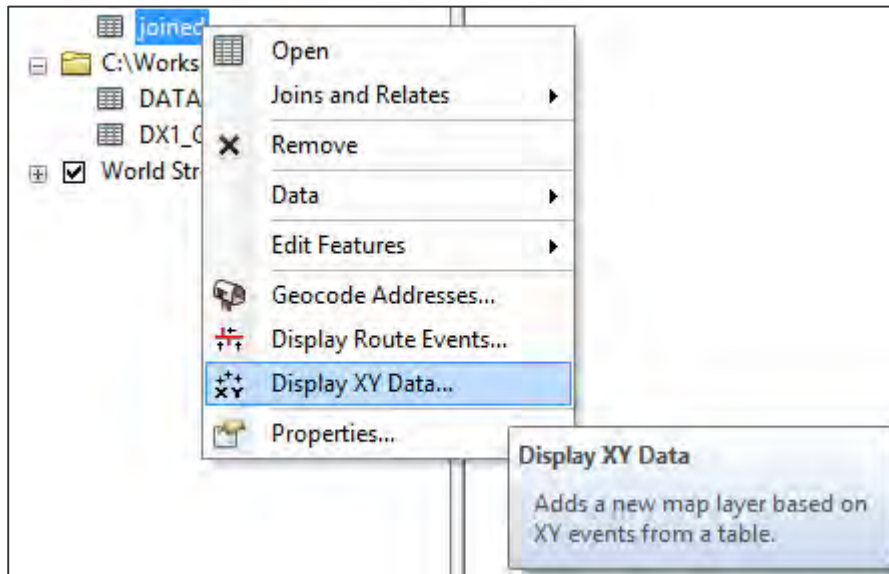


Join

	CYPCID	DOB	GENDER	POSTAL_CODE	DX_DATE	DX_AGE	DX1_GRP	CHEMO	DX1_GRP	Diagnosis_Group	FID	POSTALCODE	LONGITUDE	LATITUDE	DAUID
▶	1	1999-03-20	M	T1A2L4	2004-03-20	5	11	1	11	Hepatic tumors	5513	T1A2L4	-110.683259	50.029562	48010199
	2	1997-04-18	F	T2T3E5	2005-04-18	8	6	0	6	Renal tumors	18534	T2T3E5	-114.084688	51.02591	48060617
	3	2004-08-10	M	T2T3P9	2015-08-10	11	9	0	9	Other malignant epithelial neoplasms and malignant melanomas	18191	T2T3P9	-114.092444	51.023647	48060619
	4	1999-03-30	F	T1A4K1	2002-03-30	3	5	0	5	Soft tissue and other extrasosseous sarcomas	5680	T1A4K1	-110.692985	50.032527	48010194
	5	1997-05-31	F	T6B0W7	2008-05-31	11	5	1	5	Soft tissue and other extrasosseous sarcomas	63082	T6B0W7	-113.436916	53.529479	48110821
	6	1998-08-12	M	T9H2V4	2012-08-12	14	12	0	12	Other and unspecified malignant neoplasms	87789	T9H2V4	-111.364116	56.698394	48160122

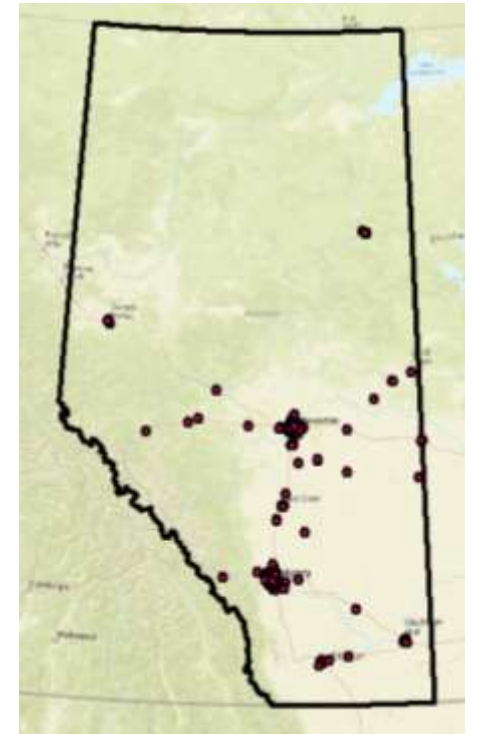
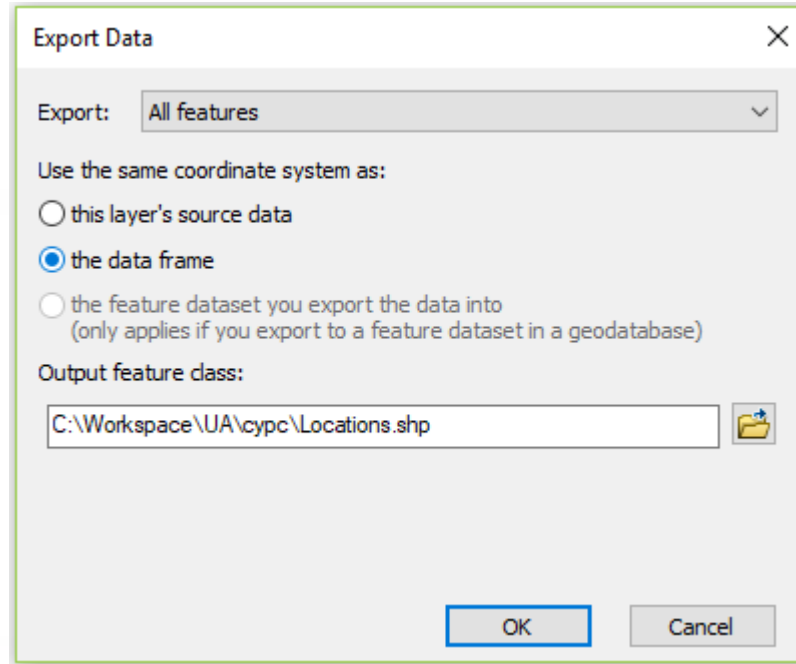
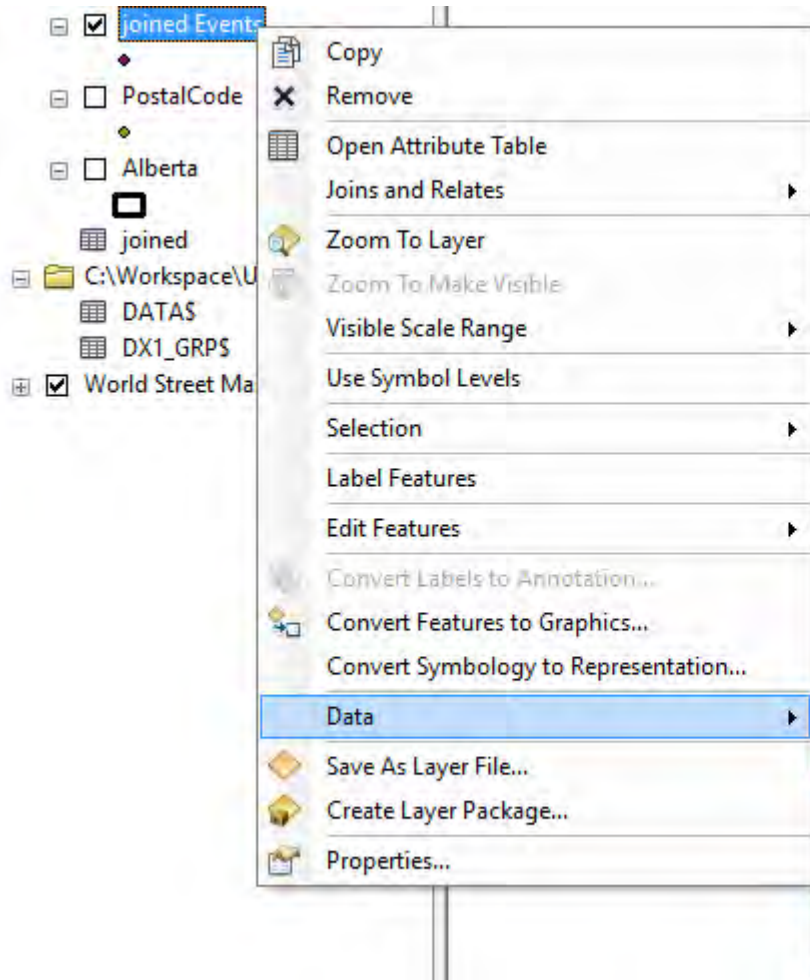


Export table



Extremely important  
to set the correct  
coordinate system

Display XY Events



Locations.shp is a true spatial data file

Export data

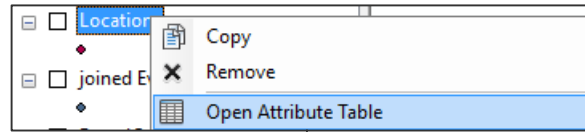
# Sample analyses

- **Summary Statistics**
- **Overlay**
- **Proximity**



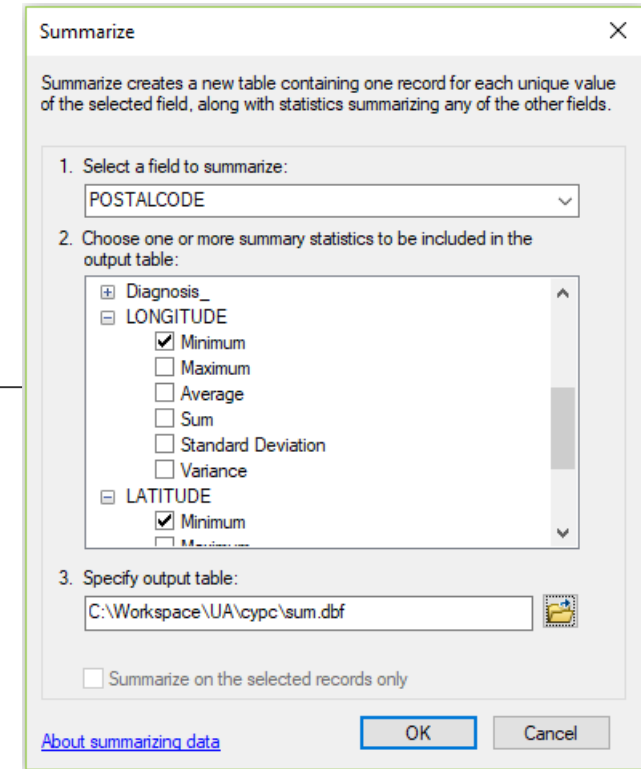


# Summarize (counts)



POSTALCODE	LONGITUDE	LATITUDE	DAVID
T1A2L4			199
T2T3E5			617
T2T3P9			619
T1A4K1			194
T6B0W7			821
T9H2V4			
T6H3B1			
T3M0B3			
T1X0T1			
T2K6E2			
T2A2E5			
T6H0M7			
T2H2V9			
T8V3J5			

**Summarize**  
Create a summary table grouped by the values in this field. The dialog that appears lets you choose whether all the records will be summarized or just the selected records.

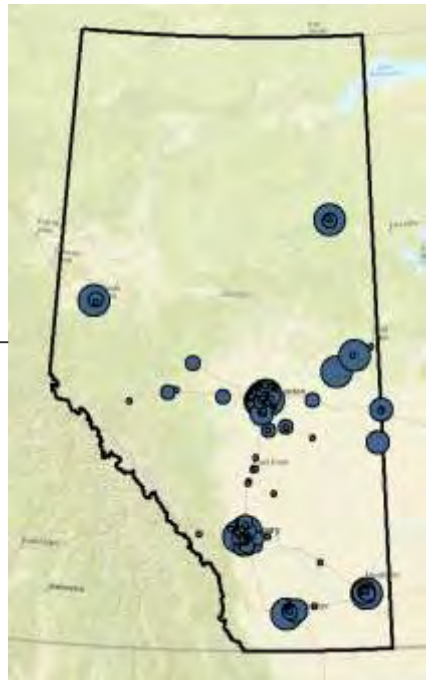
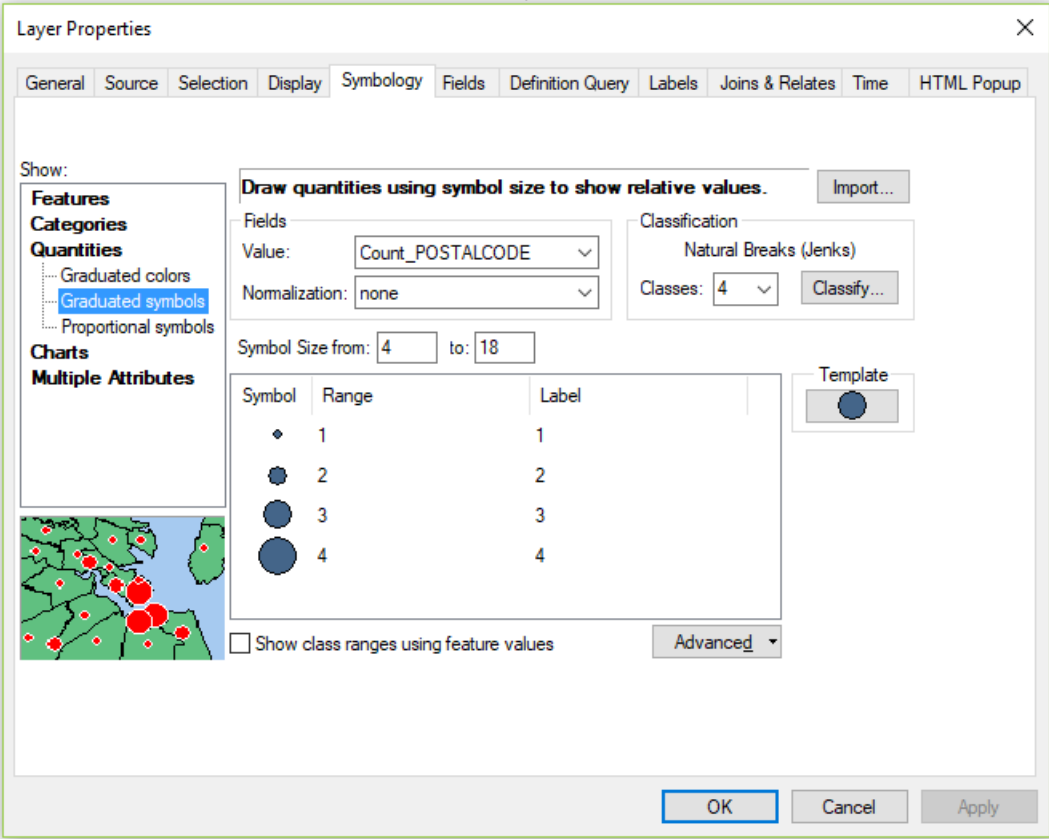
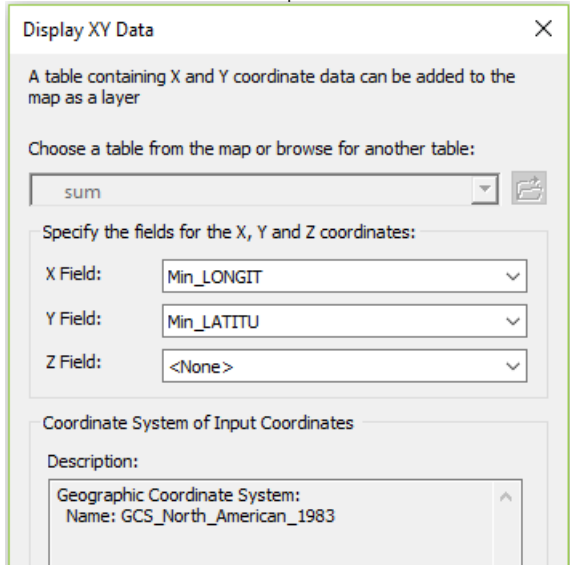
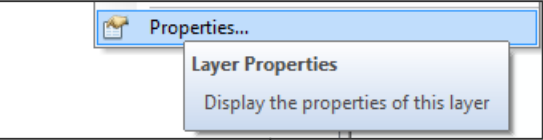
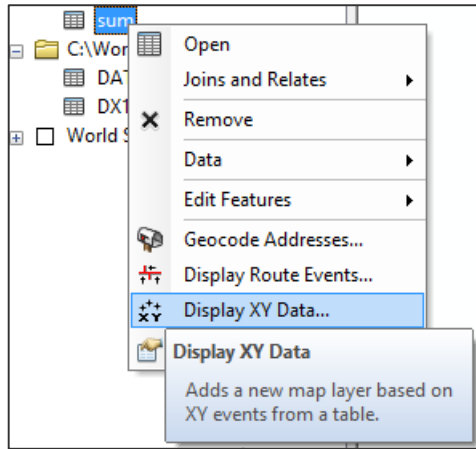


OID	POSTALCODE	Count_POSTALCODE	Minimum_LONGITUDE	Minimum_LATITUDE
0	T0A3A0	4	-111.29564	53.991485
1	T0B0H6	1	-112.062921	52.812014
2	T0B0V0	3	-110.141934	52.693778
3	T0E0B9	2	-114.737235	53.556229
4	T0M0Z0	1	-113.227761	51.828433
5	T1A1B1	1	110.678976	50.029482

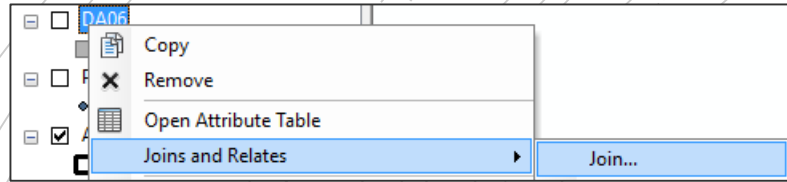
(0 out of 200 Selected)

Locations sum





Visualize point counts



# Overlay (point in polygon)

**Join Data**

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

Locations

2. You are joining: Points to Polygons

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

Each polygon will be given a summary of the numeric attributes of the points that fall inside it, and a count field showing how many points fall inside it.

How do you want the attributes to be summarized?

Average  Minimum  Standard Deviation

Sum  Maximum  Variance

Each polygon will be given all the attributes of the point that is closest to its boundary, and a distance field showing how close the point is (in the units of the target layer).

Note: A point falling inside a polygon is treated as being closest to the polygon, (i.e. a distance of 0).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

C:\Workspace\UA\cypc\overlay.shp

[About joining data](#)

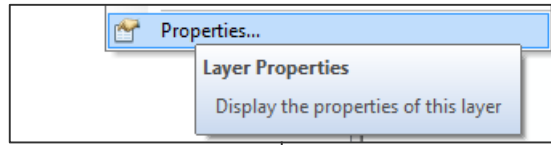
Table

overlay

FID	Shape	Locations_FID	OBJECTID	DAUID	Count_
4312	Polygon	4312	45907	48111965	0
4313	Polygon	4313	45908	48111966	1
4314	Polygon	4314	45909	48111967	0
4315	Polygon	4315	45910	48111968	0
4316	Polygon	4316	45911	48111969	0
4317	Polygon	4317	45912	48111970	0
4318	Polygon	4318	45913	48111971	0
4319	Polygon	4319	45914	48111972	0
4320	Polygon	4320	45915	48111973	3
4321	Polygon	4321	45916	48111974	0
4322	Polygon	4322	45917	48111975	0

(1 out of 5357 Selected)





Layer Properties

General Source Selection Display Symbology Fields Definition Query Labels Joins & Relates Time HTML Popup

Show:

- Features
- Categories
- Quantities
  - Graduated colors
  - Graduated symbols
  - Proportional symbols
  - Dot density
- Charts
- Multiple Attributes

**Draw quantities using color to show values.** Import...

Fields: Value: Count\_ Classification: Natural Breaks (Jenks)

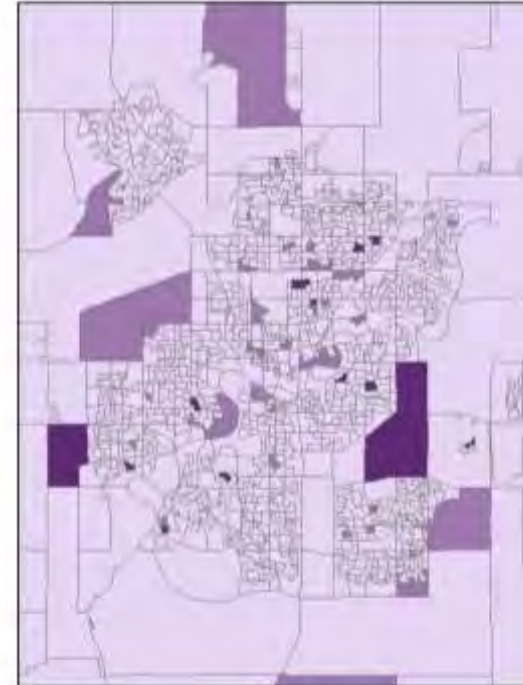
Normalization: none Classes: 3 Classify...

Color Ramp: [Color Ramp]

Symbol	Range	Label
[Light Purple]	0	0
[Medium Purple]	1 - 2	1 - 2
[Dark Purple]	3 - 6	3 - 6

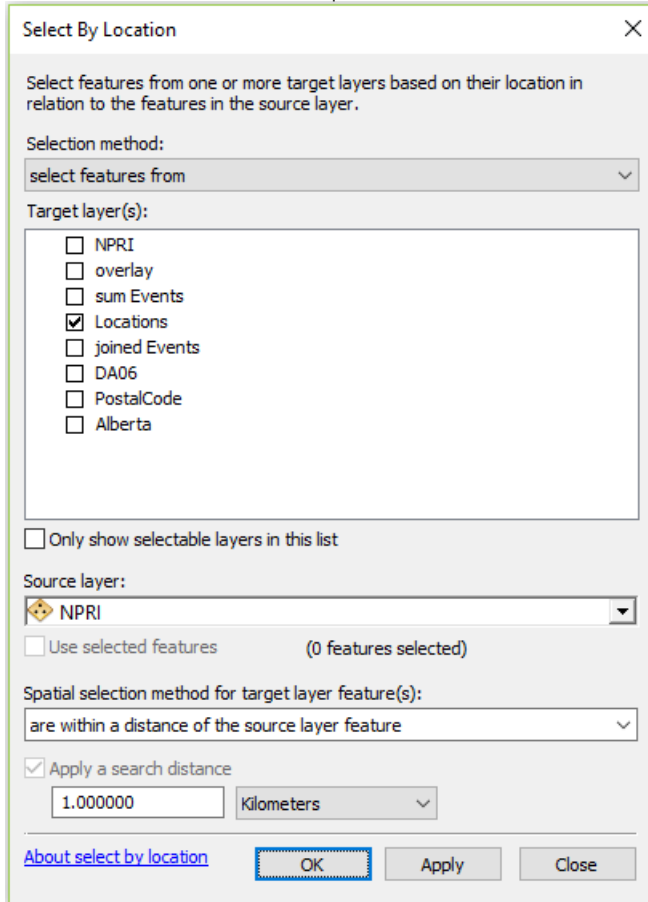
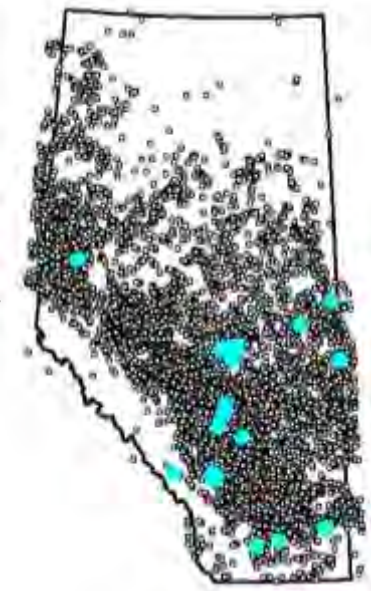
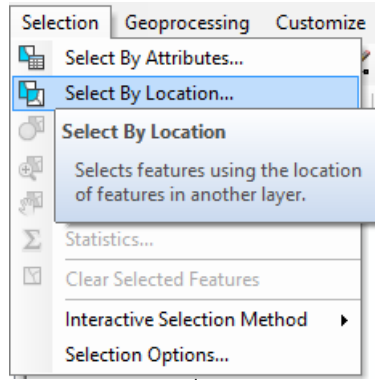
Show class ranges using feature values Advanced

OK Cancel Apply



Visualize area counts

# Proximity (closest features)

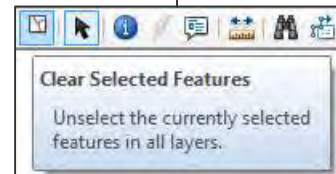


Table

Locations

CYPCID	DOB	GENDER	POSTAL_COD	DX_DATE
38	1995-01-08	M	T8A2Y2	2003-01-08
39	2008-08-17	M	T2G1C9	2012-08-17
40	2000-11-28	M	T3B1M5	2014-11-28
41	1999-02-15	M	T9V3A4	2010-02-15
42	2005-01-21	F	T5H1N2	2010-01-21
43	2008-01-30	F	T5R3X3	2012-01-30
44	1995-11-27	M	T3A2C9	2003-11-27
45	2002-05-27	M	T7X4M6	2006-05-27
46	1994-10-12	M	T3E5L9	2002-10-12

(97 out of 333 Selected)





**Join Data**

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

NPRI

2. You are joining: Points to Points

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

Each point will be given a summary of the numeric attributes of the points in the layer being joined that are closest to it, and a count field showing how many points are closest to it.

How do you want the attributes to be summarized?

Average  Minimum  Standard Deviation  
 Sum  Maximum  Variance

Each point will be given all the attributes of the point in the layer being joined that is closest to it, and a distance field showing how close that point is (in the units of the target layer).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

C:\Workspace\UA\cypc\proximity.shp

[About joining data](#) OK Cancel

Table

proximity

FID	Shape	CYPCID	Diagnosis_	NPRI_ID	COMP_NAME	Distance
141	Point	281	Malignant bone tumors	0000006580	University of Lethbridge	2338.0320
142	Point	330	Leukemias, myeloproliferative dis	0000006580	University of Lethbridge	2338.0320
143	Point	54	Neuroblastoma and other periphe	0000006607	Husky Oil Operations Limited	856.4980
144	Point	38	Soft tissue and other extraosseo	0000006660	Shell Canada Products	4252.2690
145	Point	12	Retinoblastoma	0000006678	University of Alberta Heating Plant	3714.6954
146	Point	20	Soft tissue and other extraosseo	0000006678	University of Alberta Heating Plant	3170.0500
147	Point	43	CNS and miscellaneous intracrani	0000006678	University of Alberta Heating Plant	3191.1772
148	Point	77	Retinoblastoma	0000006678	University of Alberta Heating Plant	780.7193
149	Point	121	Other malignant epithelial neoplas	0000006678	University of Alberta Heating Plant	1086.8422
150	Point	251	Hepatic tumors	0000006678	University of Alberta Heating Plant	3714.6954

(0 out of 333 Selected)

Sort Ascending

Sort Descending

Advanced Sorting...

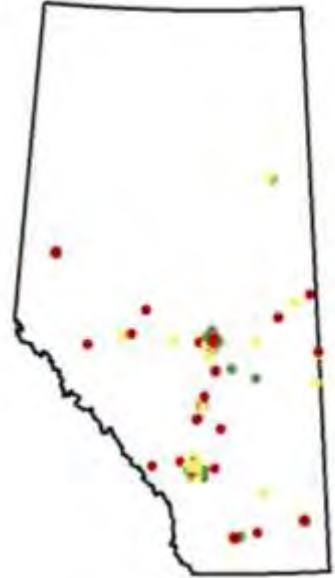
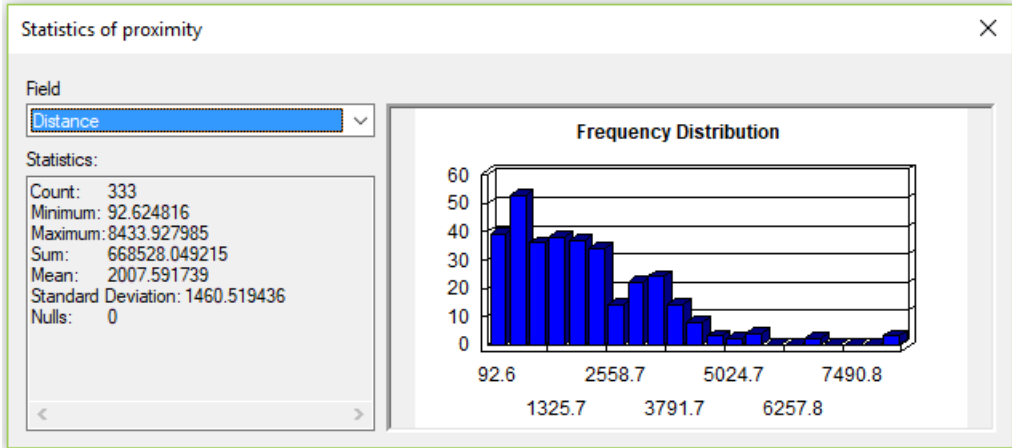
Summarize...

**Statistics...**

Field Calculator...

**Statistics**

Generates a report of statistics for the selected values in this numeric field. This command is disabled if this field is not numeric. If any of the records in the table are currently selected, statistics will only be generated for the selected records.



Visualize distances

# GIS Demo



More  
learning...

- GIS and Public Health (Cromley and McLafferty)  
<https://www.library.ualberta.ca/catalog/5440524>
- GIS Tutorial for Health (Kurland and Gorr)  
<https://www.library.ualberta.ca/catalog/7769694>
- CDC Training in GIS and Public Health  
<https://www.cdc.gov/gis/gis-training.htm>
- Unlimited Esri Training for the ArcGIS platform  
<https://www.esri.com/training/unlimited-esri-training/>
- ...and a resource – SAS Bridge for Esri  
[https://www.sas.com/en\\_ca/software/bridge-for-esri.html](https://www.sas.com/en_ca/software/bridge-for-esri.html)



# Recap

- Overview of GIS
  - A system of 5 important components
- Basemaps and map projections
  - Get started with context and coordinate systems
- Geolocation
  - Locating our tabular data by postal codes
- Analyses: statistics, overlay, and proximity
  - Where we have higher/lower counts and distances to other geographic features
- Other learning references



## Acknowledgements:

- Dr. Alvaro Osornio-Vargas
- Jesus Serrano-Lomelin
- The DoMiNO Project Team
- Randy Barber and the CYP-C Program

Thank you!



Cancer  
Research  
Society



UNIVERSITY OF ALBERTA  
FACULTY OF MEDICINE & DENTISTRY  
Department of Pediatrics



UNIVERSITY OF ALBERTA  
FACULTY OF SCIENCE  
Department of Computing Science



UNIVERSITY OF ALBERTA  
FACULTY OF SCIENCE  
Department of Earth & Atmospheric Sciences